GROUP-WISE MDL BASED REGISTRATION OF SMALL ANIMALS IN VIDEO SEQUENCES

Yuan Han, Georg Langs, Nikos Paragios

Laboratoire de Mathématiques Appliquées aux Systèmes, Ecole Centrale de Paris,
Equipe GALEN, INRIA Saclay, Ile-de-France
Grande Voie des Vignes, 92 295 Chatenay-Malabry, France
www.mas.ecp.fr

ABSTRACT

We propose a method for the non-rigid tracking of small animals in video sequences. Based on an image sequence showing the animal, first a sub-set of images with coherent pose is chosen automatically. Then a robust rigid registration determines the coarse animal body pose in the set of frames and a subsequent non-rigid group-wise registration using the minimum description length principle learns a deformation model of the animal. It obtains a non-rigid mapping between the individual positions and shapes of the animal during the sequence, based on discrete sets of landmark candidates and according texture features. This is of high relevance in small animal research to integrate signals from multiple frames, and cannot be achieved by standard continuous registration methods. We report first experimental results on video sequences of a rat.

Index Terms— Nuclear imaging, SIFT, RANSAC, Non-Rigid Registration, Group-Wise Registration, MDL, Active Appearance Models

1. INTRODUCTION

Non-rigid registration is an active area of research. In medical imaging most registration approaches address the continuous mapping between data showing anatomical structures. In these cases the deformation is usually constraint by anatomy, and the assumption of continuous deformation fields is valid. Examples for group-wise registration methods are [1] where a one-to-many non-rigid continuous registration is proposed, or [2] where the statistical properties of dense deformation fields are used to perform the registration of multiple examples.

In small animal research the monitoring and the integration of signals stemming from multiple frames to improve signal quality [3], is an important and open problem. In particular the tracking of animals over long periods of time is difficult. Continuous registration techniques can not be applied on such video sequences, since the moving animal causes discontinuous displacements, partially missing data due to occlusions, and exhibits a shape and pose variation that is beyond the range of typical anatomical structures.

We present a non-rigid registration approach, that establishes a mapping between the body of an animal at different time points. The registration is based on a robust estimation of the animal pose by means of interest points, and local texture descriptors. The non-rigid registration of the deforming animal body is formulated as a model building problem. It is addressed by a minimum description length principle based criterion, that accounts for the systematic deformation - as opposed to purely elastic - of animals at different time points. Direct applications of the method are behavioral studies of small animals, or the fusion of image information, e.g. nuclear imaging data, where the signal to noise ratio makes an integration over a long time period necessary.

This work has been partially supported from the Region Île-de-France.

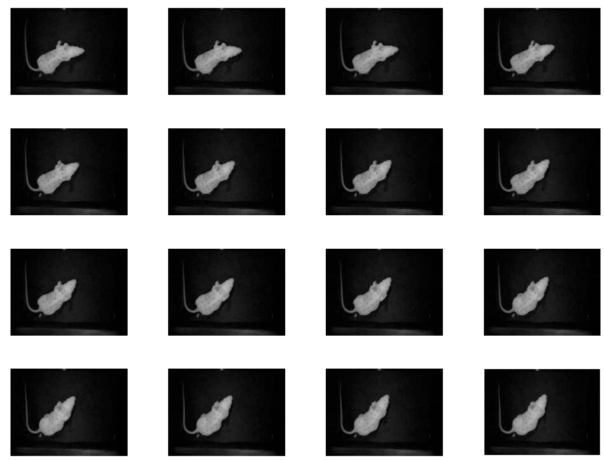

Fig. 1. Frames of a typical sequence showing a rat.

The remainder of the paper is structured as follows: in Sec. 2 we outline the method, while the two main parts of the approach, robust pose estimation, and group-wise non-rigid registration are explained in detail in Sec. 3 and Sec, 4, respectively. In Sec. 5 experimental results are reported, and a final discussion is given in Sec. 6.

2. METHODOLOGY

A small animal is monitored with a static camera. It acquires frames at fixed time intervals that are used for the registration. In Fig. 1 frames of a sequence of a rat are shown. The automatic registration method is roughly divided into three steps:

A coarse registration of the animal body is performed. The rigid registration is based on feature points extracted on the animal body and according local image descriptors. Homographies between the frames are calculated by a RANSAC [4] procedure. Then, frames in which the animal exhibits a coherent pose are detected in the image sequence. This is necessary to avoid frames that cannot be registered e.g. a rat standing vertically during a short time period of the sequence filmed from atop a cage.

In the last step the non-rigid deformation of the animal body

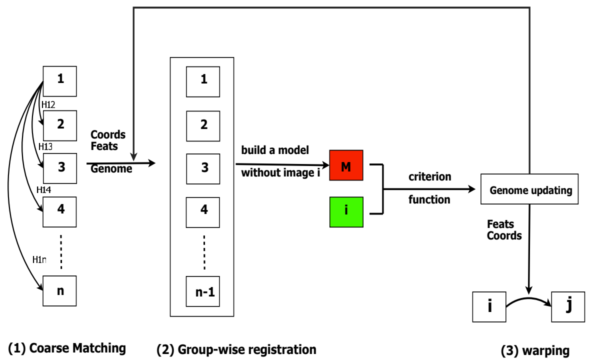

Fig. 2. Schematic overview of the algorithm.

is modeled automatically in the chosen sub-sequence. A minimum description length based optimization learns a model of shape variation, while at the same time establishing a non-rigid mapping between the different instances of the animal body. In Fig. 2 a schematic overview of the algorithm is depicted.

3. RIGID REGISTRATION AND COHERENT SUB-SEQUENCES

As a first step background subtraction is performed on the image sequence. Since in a controlled illumination situation the background (e.g. the floor of the cage) remains constant, while the animal moves, it can be modeled by the median value (or minimum in the case of black floors) of each pixel. However, other background modeling frame works like adaptive background models can by applied, if the algorithm has to account for changing illumination. After the background modeling each pixel can be classified into background or foreground, and the remaining calculation is restricted to foreground regions.

On the foreground regions SIFT features [5] are extracted for all frames, and a homography between the subsequent frames is estimated in a robust fashion using RANSAC. Please refer to [6] for a detailed discussion of homography estimation. SIFT features determine interest points and according descriptors that can be detected repeatably. The local texture description is based on gradient histograms. They have proven reliable since they are scale and rotation invariant, the latter being of particular importance in a small animal tracking scenario. No artificial external markers like e.g. infrared markers are required.

To robustly register the animal body in two frames, a pair-wise matching of the interest points based on the SIFT features in two frames $\mathbf{I}_i$ and $\mathbf{I}_j$ is performed. This results in pairs $(x_{i,k}, x_{j,k})$ of corresponding points in the two images. From this set of correspondences sub-sets of 4 points are chosen randomly, and a homography H_{ij} between images $\mathbf{I}_i$ and $\mathbf{I}_j$ is estimated. The remaining points $x_{i,l}$ in image $\mathbf{I}_i$ are transformed to positions $x'_{j,l}$ in $\mathbf{I}_j$ according to H_{ij}, and the registration error to the pair-wise match $x'_{j,l} - x_{j,l}$ is calculated. The number of *inliers*, with a registration error below a certain tolerance threshold, determines a notion of reliability of the homography. RANSAC chooses the homography H^*_{ij} with the highest number of inliers. The resulting estimates are robust against a certain amount of ambiguous matches caused by change of rat posture.

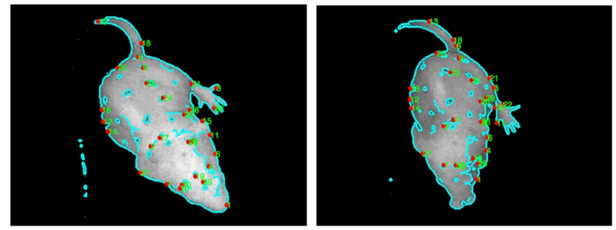

Fig. 3. Initial correspondences after rigid-registration.

We calculate homographies between subsequent images I_i and I_{i+1} with $i = 1, \ldots, n-1$. This results in a sequence of homographies from $H_{1,2} \ldots H_{i,i+1} \ldots H_{n-1n}$. The homography, $H_{1,k}$, from the reference image, I_1, to the image, I_k, is the concatenation of the consecutive homographies from H_{12}, H_{23} ...until H_{k-1k}. As two neighboring images are generally quite similar, in most cases the consecutive homographies H_{ii+1} constructed are stable.

Since the sequence is expected to contain frames, that cannot be registered without considerable distortions, we choose a sub-sequence with coherent animal behavior. The selection procedure is based on the number of inliers during homography estimation. A good strategy was to choose sub-sequences, where the number of inliers in pairwise homographies is above a certain threshold.

4. GROUP-WISE NON RIGID REGISTRATION

After a coarse robust rigid registration of the animal body a fine non-rigid registration is performed. Correspondences are initialized by the homographies $H_{i,j}$ between individual frames. Then the non-rigid group-wise registration is performed based on a method proposed in [7].

4.1. Choice of features and interest points

Both the rigid and no-rigid registration are based on finite sets of interest points, and corresponding local texture features. We compared the suitability of two different interest point detectors. For both gradient histograms were calculated to describe the local texture.

1. Difference of Gaussian (DoG) and gradient histograms: the standard SIFT [5] approach can be employed to find point correspondences. This method gives best results for the robust homography estimation between images, since SIFT features are stable, and invariant against rotation and scale changes. The drawback of SIFT in this scenario, is that the DoG interest point detector neglects a considerable amount of information on the animal body valuable for the determination of fine non-rigid displacements. The two most salient features are the contour of the animal, and in case of markers - e.g. a color dot pattern on the animal - their positions. In addition DoG points are not stable if curvature changes occur. In the case of a bending animal this results in a shifting of interest point positions. For the coarse robust rigid registration, the amount is not relevant, but it would decrease the quality of a non-rigid mapping considerably.

2. Canny edges and gradient histograms: To overcome the limitation of DoG points, Canny edge points can be used. With this method, boundary points are chosen as interest points by a canny detector. For each interest point detected, the orientation is calculated according to the gradient at its position. To describe the texture gradient histograms are then calculated analogously to the SIFT features, with a fixed scale. The fixed scale assumption is valid, if we can assume constant distance between the animal and the camera.

4.2. Group-wise registration

The group-wise registration is based on an approach proposed in [7]. The interest points in the images serve as landmark candidates. Initial correspondences on the frames $\mathbf{I}_i$, $i = 1, 2, \ldots, n$ for a set of k landmarks are established by pairwise propagation of the landmark position in $\mathbf{I}_1$ to $\mathbf{I}_i$ based on the according homography $H_{1,i}$ and the search for the closest interest point. This results in correspondences for k landmarks $\{l_1, \ldots, l_k\}$, which are encoded in a $k \times n$ matrix $\mathbf{G}$. Each column represents an image, and the entry $\mathbf{G}_{ji} \in \{1, \ldots, m_i\}$ with $j \in \{1, \ldots, k\}$ is the index of the interest point in image $\mathbf{I}_i$, at which the landmark l_j is positioned.

Each point (i, q) with $q \in \{1, \ldots, m_i\}$ is assigned its coordinate information $\mathbf{p}(i, q)$ and local features $\mathbf{f}(i, q)$ (e.g. SIFT, steerable filters). By assigning $\mathbf{G}_{ji} = q$ the landmark l_j in image $\mathbf{I}_i$ has position $\mathbf{p}_{ij} = \mathbf{p}(i, q)$ and feature vector $\mathbf{f}_{ij} = \mathbf{f}(i, q)$. Starting from these correspondences group-wise registration is performed by minimizing a criterion function that captures the compactness of the appearance model comprising the variation of landmark positions and local texture variation at the landmark positions in the frames of the sequence. In Fig. 3 initial correspondences in the set of interest points are depicted for two frames.

The criterion function is based on the minimum description length principle: the cost function is given as $\mathcal{C} = \mathcal{C}_S + \mathcal{C}_T + \alpha(t)\mathcal{C}_E$ as described in paper [7]. It couples the non-rigid group-wise registration with the building of a model, that describes the variation of shape and texture in the examples. Since we can expect, that a moving animal exhibits a certain amount of systematic deformation behavior - i.e. points cannot move independently - we can transfer the group-wise registration task to a model building task. In contrast to purely elasticity constraint registration it allows for the capturing of systematic and heterogeneous deformation patterns. The model should capture the shape and texture variation in the image sequence in the most compact manner [8]. An optimal model should minimize the cost L of communicating the model $\mathcal{M}$ itself and the data D (i.e. the landmark positions) encoded with the model: $L(D, \mathcal{M}) = L(\mathcal{M}) + L(D|\mathcal{M})$, where $L(\mathcal{M})$ is the cost of communicating the shape model, $L(D|\mathcal{M})$ is the cost of the shape data encoded with help if the model, and $\mathcal{R}$ is a penalty for the residual error not captured by the model.

The criterion function consists of the description lengths for the statistical shape model capturing the deformation of the landmarks $\mathcal{C}_S$, a texture variation model that represents the variation of local texture for each landmark in the image set $\mathcal{C}_T$, and an elasticity term, that serves as a regularization during the initial phase of the registration $\mathcal{C}_E$.

The criterion is minimized by iteratively selecting a random frame k, and building a model from the remaining frames. This model is fitted to the current status of correspondences in frame k. Based on the fitted model, new landmarks in the sample image are chosen from the set of interest points, and the positions of the landmarks in this frame are updated accordingly. The new positions are encoded in the matrix $\mathbf{G}_{ji}$ by changing the corresponding entries. A schematic overview is given in Fig. 2. This optimization results in landmarks, and according positions in each frame. Since no reference manifold, but only a discrete set of points is used for the landmark position estimation, the approach can deal with topolgy changes, and partially missing data. This is relevant since the movement of the animal can render parts invisible in some frames.

Method	(1)	(2)	(3)
d_1	26.0893	6.4577	9.2535
d_2	32.4884	10.7660	20.1352
d_3	18.5289	7.3800	6.4847
d_4	28.8441	6.0174	16.7530
d_5	20.7635	7.4360	5.1843
d	25.3428	7.6114	11.5621

Table 1. Registration error (pixels) for three different methods.

4.3. Deformation fields on the animal body

After the registration has converged a set of landmark positions are known in all images. They define a deformation field on the body of the animal. It can be interpolated by thin plate splines, or piecewise affine transformation. This yields a dense mapping between the regions of the animal in individual frames, and can be used for the integration of signals (e.g. nuclear imaging).

5. EXPERIMENTAL RESULTS

We have performed initial experiments on 5 sequences depicting a rat in a cage, filmed vertically from atop. Each sequence consists of fifteen to twenty images.We evaluate the accuracy of the coarse rigid registration and the fine non-rigid registration. For 8 points, corresponding positions were annotated manually. To obtain a quantitative assessment of the registration accuracy, the positions of these points were propagated across the entire sequence based on the correspondences resulting from registration.

For the coarse registration we compare three different strategies of extracting interest points and registering the animal body:

(1) Pair-wise point matching with DoG/SIFT: interest points and descriptors are obtained by the standard SIFT approach. Landmarks are chosen randomly from the key-points in the image I_1, and are matche based on the local texture features.

(2) Robust homography estimation and non-rigid registration based on DoG/SIFT: The key-points are first obtained by SIFT and a homography is estimated from the set of correspondences by RANSAC. Lastly, the landmark correspondences are determined by applying the concatenated homographies and matching to the closest interest points in each image.

(3) Robust homography estimation using DoG/SIFT correspondences, non-rigid registration based on canny points and gradient histograms: Analogous to (2) but Canny edge points are used for the propagation of correspondences after homography estimation.

In each image sequence, we mark 8 reference points manually in I_1 and the corresponding 8 points in I_5,$I_{10}I_{15}$ are calculated automatically. Throughout the experiment, we obtain, for each sequence, 8×3 manually selected ground truth points and an equal number of automatically calculated points and calculate the mean distance d_i between the manual points and the calculated points in each sequence. Finally, the average of these mean distances $\hat{d}$ is evaluated. Tab. 1 shows the quantitative performance of the different coarse registration methods. The robust homography estimation leads to a significantly improved accuracy over method (1). The initial correspondences of method (2) and (3) are in the same range.

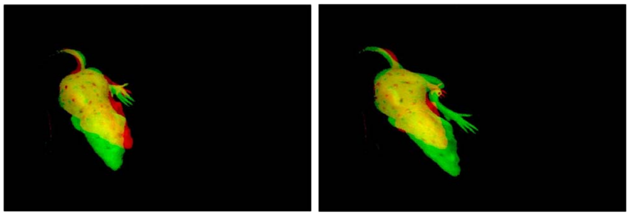

Fig. 4. Non-rigid registration result based on DoG points and RANSAC initialization. Note that the registration was restricted to the body, i.e., the head was not registered.

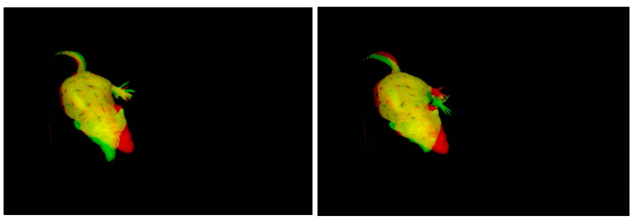

Fig. 5. Non-rigid registration result based on Canny points and RANSAC initialization. Note that the registration was restricted to the body, i.e., the head was not registered.

Based on the previous results, the initial approximation of correspondences for the group-wise registration can be appropriately obtained through the coarse methods (2) or (3). To estimate the result of group-wise registration, the results of a dense mapping of the reference frame animal body to the remaining frames is performed. The mapping is defined by the landmark correspondences resulting from registration. In Figs. 4 and 5 overlays of the results of non-rigid registration with the actual frame are depicted for two example frames. The red shapes are the visualized image in the sequence and the green shapes are the mapped image from I_1. In Fig. 4 the non-rigid registration was based on method (2), while in Fig. 5 the result of method (3) is depicted. The comparison indicates that the precision of the fine registration is restricted to the level of accuracy and density of the interest point locations. The results show that Canny points are better suited to serve as landmark candidates of the non-rigid registration. However, inaccuracies remain, and are subject of ongoing research.

6. CONCLUSION

This paper proposes a method for the non-rigid registration of video sequences showing small animals. From a set of frames first a suitable sub-sequence is selected, and robust rigid registration based on interest points and local texture descriptors is performed. This serves as an initialization for a non-rigid group-wise registration of the animal body. It accounts for systematic deformation behavior of the animal to improve the registration result. The resulting correspondences establish a deformation field on the body, that allows for the integration of signals over multiple frames. This is highly relevant in nuclear imaging modalities, since the short exposure time possible with a moving animal causes bad signal to noise ratio. We have compared the suitability of different interest points for the sparse group-wise registration, and first experiments show that they have a considerable influence on the result. Future work will concentrate on the improvement of the non-rigid registration, and the application to nuclear imaging [9].

7. REFERENCES

[1] D. Rueckert, A.F. Frangi, and J.A. Schnabel, "Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration," *IEEE TMI*, vol. 22, no. 8, pp. 1014–1025, 2003.

[2] T.F. Cootes, C.J. Twining, V. Petrović, and C.J. Taylor, "Groupwise construction of appearance models using piecewise affine deformations," in *Proc. of BMVC'05*, 2005.

[3] J. S. Goddard, S. S. Gleason, M. J. Paulus, S. Majewski, V. Popov, M. Smith, A. Weisenberger, B. Welch, and R. Wojcik, "Pose measurement and tracking system for motion-correction of unrestrained small animal pet/spect imaging," in *Proc. IEEE Medical Imaging Conference*, 2003.

[4] Martin A.Fischler and Robert C.Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[5] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[6] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[7] G. Langs, R. Donner, P. Peloschek, and H. Bischof, "Robust autonomous model learning from 2D and 3D data sets," in *Proc. MICCAI'07*, 2007, vol. LNCS 4791, pp. 968–976.

[8] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[9] A.G. Weisenberger, S.S. Gleason, J. Goddard, B. Kross, S. Majewski, S.R. Meikle, M.J. Paulus, M. Pomper, V. Popov, M.F. Smith, B.L. Welch, and R. Wojcik, "A restraint-free small animal spect imaging system with motion tracking," *IEEE Transactions on Nuclear Science*, vol. 52, no. 3, pp. 638– 644, 2005.

[10] R.H. Davies, C.J. Twining, T.F. Cootes, J.C. Waterton, and C.J. Taylor, "A minimum description length approach to statistical shape modeling," *IEEE TMI*, vol. 21, no. 5, pp. 525–537, May 2002.

[11] B. Ma, Ramkrishnan Narayanan, Hyunjin Park, Alfred O. Hero, Peyton H. Bland, and Charles R. Meyer, "Comparing pairwise and simultaneous joint registrations of decorrelating interval exams using entropic graphs.," in *Inf Process Med Imaging*, Nico Karssemeijer and Boudewijn P. F. Lelieveldt, Eds. July 2007, vol. LNCS 4584, pp. 270–282, Springer.

[12] D. Rueckert, A. F. Frangi, and J. A. Schnabel, "Automatic construction of 3d statistical deformation models using non-rigid registration," in *Proc. MICCAI'01*, W.J. Niessen and M.A. Viergever, Eds., Berlin, Germany, 2001, vol. 2208 of *Lecture Notes in COmputer Science*, pp. 77–84, Springer Verlag.

[13] Zambanini S., Langs G., Sablatnig R., and Maier H., "Automatic robust registration of cutaneous hemangiomas for follow-up examinations," 31st AAPR/OAGM Workshop, 2007, pp. 121–128.